

# Simulated Totals for Hispanic National Origin Groups by State, Place, County, and Census Tract

Technical Documentation  
Issued April 2003

---

## CONTENTS

### CHAPTERS

1	Abstract . . . . .	1-1
2	Overview: Why a File Showing Simulated Hispanic Groups . . . . .	2-1
3	Procedure for Obtaining Simulated Totals . . . . .	3-1
4	Data File Layouts . . . . .	4-1
5	Accuracy of the Estimates . . . . .	5-1
6	User Updates . . . . .	6-1

# Chapter 1.

## Abstract

---

	Page
Citation . . . . .	1-1
Type of File . . . . .	1-1
Subject Content . . . . .	1-1
Geographic Content . . . . .	1-1
Technical Description of Data Files . . . . .	1-2

### CITATION

Simulated Totals for Hispanic National Origin Groups by State, Place, County, and Census Tract, prepared by the U.S. Census Bureau, 2003.

### TYPE OF FILE

Summary statistics.

### SUBJECT CONTENT

The File provides totals for detailed Hispanic groups from Census 2000 and from the results of a simulation analysis.

The simulated totals presented in this report are illustrative and do not replace the official Census 2000 totals for these groups. The totals from Census 2000 are based on sample estimates from the Spanish/Hispanic/Latino question on the long form and are the official figures. The simulated totals provided in this paper do not fully reflect self-identification. Instead, they include: (1) the responses of those who self-identified into a specific group and (2) the responses of those who reported a general term and were tabulated into a specific group using responses to the place of birth or ancestry questions. In other words, respondents who provided a more general Hispanic response might, or might not, have provided a more specific response if the question wording or format had been different. It is impossible to predict with certainty how they would have responded to the Hispanic origin question had it been phrased differently. Given the increasingly frequent use of the more general terms "Hispanic" and "Latino" in the media, it would not be surprising if increasing numbers of people, especially those whose ancestors had come from Spanish-speaking countries two or more generations ago, identified with these terms rather than specific national origins.

For more information, see Working Paper #72, "Analysis of General Hispanic Responses in Census 2000," included with this data product

Questions about the subject content in this file should be directed to: Population Division, U. S. Census Bureau, Washington DC 20233-8800. Phone: 301-763-2422.

### GEOGRAPHIC CONTENT

Estimates from the Census 2000 Spanish/Hispanic/Latino question and simulated totals for each Hispanic group are provided for the Nation, states, places, counties and census tracts within counties. The Federal Information Processing Standards (FIPS) code is used to identify each state and county.

## TECHNICAL DESCRIPTION OF DATA FILES

### ASCII FILE:

Record length	441 characters
Data Set Name	simhisp1.dat (Alabama - Ohio)
Record count	61,503
Record length	441 characters
Data Set Name	simhisp2.dat (Oklahoma-Wyoming)
Record count	23,090

### COMMA DELIMITED FILE:

Record length	437 (longest)
Data Set Name	simhisp1.csv (Alabama - Ohio)
Record count	61,504
Record length	437 (longest)
Data Set Name	simhisp2.csv (Oklahoma-Wyoming)
Record count	23,091

### DBASE FILE:

Version of dBase	dBase III+
Data Set Name	simhisp1.dbf (Alabama-Ohio)
Record count	61,504
Version of dBase	dBase III+
Data Set Name	simhisp2.dbf (Oklahoma-Wyoming)
Record count	23,091

## Chapter 2.

### Overview: Why a File Showing Simulated Hispanic Groups

---

This file was produced as part of the analysis to explore issues raised by data users that the Census 2000 question on Hispanic origin influenced respondents to provide more general Hispanic responses (such as, “Spanish,” “Hispanic,” and “Latino”) instead of Hispanic national origin responses (such as “Mexican,” “Salvadoran,” and “Dominican”). The Working Paper #72, “Analysis of General Hispanic Responses in Census 2000,” which accompanies this data file, provides a detailed explanation of the methodology used to derive these simulated totals and discusses the results of research on this issue conducted in the Census Bureau and by other researchers.

Several limitations to this simulation apply. First, it assumes that the specific Hispanic origin groups can only grow larger while the Hispanic total is left unchanged. However, 2.6 million people indicated a general Hispanic response in the question on origin but did not provide an Hispanic national origin in the place of birth or ancestry questions. In these cases, we did not assume for the purpose of the simulation analysis that they were not Hispanic. Second, the simulation made no adjustment when individuals reported an Hispanic national-origin group in the question on origin but provided no other evidence of Hispanic ethnicity in the place of birth or ancestry questions. Third, the simulation made no adjustment when individuals indicated they were not Hispanic in the question on origin but did provide evidence of Hispanic ethnicity in the place of birth or ancestry questions. Fourth, people reporting mixed Hispanic ethnicities (for example, Mexican–Puerto Rican, Cuban–Salvadoran, and so on) had the same unique code (291) and were tabulated in the “Other Spanish/Hispanic/ Latino” category. The original response was not available in the file used to conduct this simulation, therefore, we could not choose one of the two Hispanic origins reported and tabulate their response in the appropriate Hispanic group.

The second and third limitations, which involve apparent inconsistencies, reflect the complexity underlying the reporting of ethnicity and highlight the problem of trying to simulate or “second guess” the self-identification of respondents using other indicators of ethnicity. Trying to develop a composite measure of Hispanic ethnicity using a combination of responses from the Hispanic origin, place of birth, and ancestry questions undermines the principle of self-identification and can lead to endless discussion about who is “Hispanic” and what is the size of the Hispanic population. In fact, the experience of using multiple indicators of Hispanic ethnicity in the 1970 census led the Census Bureau to decide that self-identification using a single question on Hispanic origin was the best method for counting this population group.

## Chapter 3.

### Procedure for Obtaining Simulated Totals

---

The simulation results presented in this working paper consisted of using the information from the place of birth and ancestry questions on the Census 2000 long (sample) form to examine possible specific Hispanic origins for people who reported a general response to the Hispanic origin question.

The process for attempting to obtain a more specific national origin consisted of four steps:

1. We selected the people who identified themselves in general Hispanic terms: Spanish, Hispanic, Latino, and so on.
2. We examined their place of birth responses to see if they reported a census-defined Hispanic country. If so, the national-origin code for that country was used. For example, if respondents reported they were “Latino” in the question on Hispanic origin and their place of birth was El Salvador, then Salvadoran was used in this simulation.
3. If the country of birth was not a census-defined Hispanic country, we examined the first ancestry reported (if any). If the first ancestry reported was an Hispanic national origin, then that specific origin was used in this simulation.
4. If the first ancestry was not Hispanic, we examined the second ancestry reported (if any). If the second ancestry reported was Hispanic, then that specific origin was used. If the second ancestry was not an Hispanic national origin or was not reported, then the original general Hispanic response remained.<sup>1</sup>

---

<sup>1</sup> This simulation also was run using responses from the ancestry question first and then using responses from the place of birth question and the results were about the same.

Simulated Totals for Hispanic National Origin Groups

## Chapter 4.

### Data File Layouts

---

#### ASCII FILE LAYOUT

<u>Character</u>	<u>Description</u>
1 - 3	Summary Level 040 = State 050 = County 140 = Tract 160 = Place
4 - 93	Area Name *
94 - 95	FIPS State Code
96 - 98	FIPS County Code
99 - 104	Tract Code
105 - 109	Place Code
110 - 117	Total Population - Census 2000
118 - 125	Total Hispanic - Census 2000
126 - 133	Total Hispanic - Based on Simulation (same as Census 2000)
134 - 139	Hispanic - Census 2000
140 - 145	Hispanic - Based on Simulation
146 - 151	Latino - Census 2000
152 - 156	Latino - Based on Simulation
157 - 162	Spanish - Census 2000
163 - 168	Spanish - Based on Simulation
169 - 175	Mexican - Census 2000
176 - 182	Mexican - Based on Simulation
183 - 189	Puerto Rican - Census 2000
190 - 196	Puerto Rican - Based on Simulation
197 - 202	Cuban - Census 2000
203 - 208	Cuban - Based on Simulation
209 - 214	Central American - Census 2000 **
215 - 220	Central American - Based on Simulation **
221 - 225	Costa Rican - Census 2000
226 - 230	Costa Rican - Based on Simulation
231 - 236	Guatemalan - Census 2000
237 - 242	Guatemalan - Based on Simulation
243 - 247	Honduran - Census 2000
248 - 252	Honduran - Based on Simulation
253 - 258	Nicaraguan - Census 2000
259 - 264	Nicaraguan - Based on Simulation
265 - 269	Panamanian - Census 2000
270 - 274	Panamanian - Based on Simulation
275 - 280	Salvadoran - Census 2000
281 - 286	Salvadoran - Based on Simulation
287 - 291	Other Central American - Census 2000
292 - 296	Other Central American - Based on Simulation

<u>Character</u>	<u>Description</u>
------------------	--------------------

297 - 302	South American - Census 2000 ***
303 - 308	South American - Based on Simulation ***
309 - 313	Argentinian - Census 2000
314 - 318	Argentinian - Based on Simulation
319 - 323	Bolivian - Census 2000
324 - 328	Bolivian - Based on Simulation
329 - 333	Chilean - Census 2000
334 - 338	Chilean - Based on Simulation
339 - 344	Colombian - Census 2000
345 - 350	Colombian - Based on Simulation
351 - 356	Ecuadorian - Census 2000
357 - 362	Ecuadorian - Based on Simulation
363 - 366	Paraguayan - Census 2000
367 - 370	Paraguayan - Based on Simulation
371 - 375	Peruvian - Census 2000
376 - 380	Peruvian - Based on Simulation
381 - 384	Uruguayan - Census 2000
385 - 388	Uruguayan - Based on Simulation
389 - 393	Venezuelan - Census 2000
394 - 398	Venezuelan - Based on Simulation
399 - 403	Other South American - Census 2000
404 - 407	Other South American - Based on Simulation
408 - 413	Dominican - Census 2000
414 - 419	Dominican - Based on Simulation
420 - 424	Spaniard - Census 2000
425 - 429	Spaniard - Based on Simulation
430 - 435	Other Hispanic - Census 2000
436 - 441	Other Hispanic - Based on Simulation

\* "Area Name" variable is left justified with trailing blanks. All other variables are right justified with leading blanks.

\*\* "Central American" is the sum of totals for the following groups: "Costa Rican," "Guatemalan," "Honduran," "Nicaraguan," "Panamanian," "Salvadoran," and "Other Central American."

\*\*\* "South American" is the sum of totals for the following groups: "Argentinian," "Bolivian," "Chilean," "Colombian," "Ecuadorian," "Paraguayan," "Peruvian," "Uruguayan," "Venezuelan," "Other South American."

CSV (Comma delimited) and DBF (Dbase) FILE LAYOUTS\*

<u>Variable name</u>	<u>Description</u>
sumlev	Summary Level 040 = State 050 = County 140 = Tract 160 = Place
namelsad	Area Name
state	FIPS State Code
county	FIPS County Code
place	Place Code
tract	Tract Code
totpop	Total Population – Census 2000
toth	Total Hispanic – Census 2000
toth_sim	Total Hispanic – Based on Simulation (same as Census 2000)
hisp	Hispanic – Census 2000
hisp_sim	Hispanic – Based on Simulation
latn	Latino – Census 2000
latn_sim	Latino – Based on Simulation
spsh	Spanish – Census 2000
spsh_sim	Spanish – Based on Simulation
mex	Mexican – Census 2000
mex_sim	Mexican – Based on Simulation
pr	Puerto Rican – Census 2000
pr_sim	Puerto Rican – Based on Simulation
cubn	Cuban – Census 2000
cubn_sim	Cuban – Based on Simulation
cam	Central American – Census 2000 **
cam_sim	Central American – Based on Simulation ***
cost	Costa Rican – Census 2000
cost_sim	Costa Rican – Based on Simulation
guat	Guatemalan – Census 2000
guat_sim	Guatemalan – Based on Simulation
hond	Honduran – Census 2000
hond_sim	Honduran – Based on Simulation
nicar	Nicaraguan – Census 2000
nicar_sim	Nicaraguan – Based on Simulation
pan	Panamanian – Census 2000
pan_sim	Panamanian – Based on Simulation
salva	Salvadoran – Census 2000
salva_sim	Salvadoran – Based on Simulation



<u>Character</u>	<u>Description</u>
ocam	Other Central American – Census 2000
ocam_sim	Other Central American – Based on Simulation
sam	South American – Census 2000 ****
sam_sim	South American – Based on Simulation *****
arg	Argentinian – Census 2000
arg_sim	Argentinian – Based on Simulation
boli	Bolivian – Census 2000
boli_sim	Bolivian – Based on Simulation
chil	Chilean – Census 2000
chil_sim	Chilean – Based on Simulation
col	Colombian – Census 2000
col_sim	Colombian – Based on Simulation
ecua	Ecuadorian – Census 2000
ecua_sim	Ecuadorian – Based on Simulation
para	Paraguayan – Census 2000
para_sim	Paraguayan – Based on Simulation
peru	Peruvian – Census 2000
peru_sim	Peruvian – Based on Simulation
urug	Uruguayan – Census 2000
urug_sim	Uruguayan – Based on Simulation
ven	Venezuelan – Census 2000
ven_sim	Venezuelan – Based on Simulation
osam	Other South American – Census 2000
osam_sim	Other South American – Based on Simulation
dom	Dominican – Census 2000
dom_sim	Dominican – Based on Simulation
sprd	Spaniard – Census 2000
sprd_sim	Spaniard – Based on Simulation
ohis	Other Hispanic – Census 2000
ohis_sim	Other Hispanic – Based on Simulation

- \* The first record of the CSV data file contains the variable names.
- \*\* “cam” is the sum of totals for the following variables: “cost,” “guat,” “hond,” “nic,” “pan,” “salva,” and “ocam.”
- \*\*\* “cam\_sim” is the sum of totals for the following variables: “cost\_sim,” “guat\_sim,” “hond\_sim,” “nic\_sim,” “pan\_sim,” “salva\_sim,” and “ocam\_sim.”
- \*\*\*\* “sam” is the sum of totals for the following variables: “arg,” “boli,” “chil,” “col,” “ecua,” “para,” “peru,” “urug,” “ven,” “osam.”
- \*\*\*\*\* “sam\_sim” is the sum of totals for the following variables: “arg\_sim,” “boli\_sim,” “chil\_sim,” “col\_sim,” “ecua\_sim,” “para\_sim,” “peru\_sim,” “urug\_sim,” “ven\_sim,” “osam\_sim.”

## Chapter 5.

### Accuracy of the Estimates

---

The data contained in this paper are based on the sample of households who responded to the Census 2000 long form. Nationally, approximately one out of every six housing units was included in this sample. As a result, the sample estimates may differ somewhat from the 100-percent figures that would have been obtained if all housing units, people within those housing units, and people living in group quarters had been enumerated using the same questionnaires, instructions, enumerators, and so forth. The sample estimates also differ from the values that would have been obtained from different samples of housing units, people within those housing units, and people living in group quarters. The deviation of a sample estimate from the average of all possible samples is called the sampling error.

In addition to the variability that arises from the sampling procedures, both sample data and 100-percent data are subject to nonsampling error. Nonsampling error may be introduced during any of the various complex operations used to collect and process data. Such errors may include: not enumerating every household or every person in the population, failing to obtain all required information from the respondents, obtaining incorrect or inconsistent information, and recording information incorrectly. In addition, errors can occur during the field review of the enumerators' work, during clerical handling of the census questionnaires, or during the electronic processing of the questionnaires.

Nonsampling error may affect the data in two ways: (1) errors that are introduced randomly will increase the variability of the data and, therefore, should be reflected in the standard errors; and (2) errors that tend to be consistent in one direction will bias both sample and 100-percent data in that direction. For example, if respondents consistently tend to underreport their incomes, then the resulting estimates of households or families by income category will tend to be understated for the higher income categories and overstated for the lower income categories. Such biases are not reflected in the standard errors.

While it is impossible to completely eliminate error from an operation as large and complex as the decennial census, the Census Bureau attempts to control the sources of such error during the data collection and processing operations. The primary sources of error and the programs instituted to control error in Census 2000 are described in detail in Summary File 3 Technical Documentation under Chapter 8, "Accuracy of the Data," located at [www.census.gov/prod/cen2000/doc/sf3.pdf](http://www.census.gov/prod/cen2000/doc/sf3.pdf).

All statements in this Census 2000 Brief have undergone statistical testing and all comparisons are significant at the 90-percent confidence level, unless otherwise noted. Further information on the accuracy of the data is located at [www.census.gov/prod/cen2000/doc/sf3.pdf](http://www.census.gov/prod/cen2000/doc/sf3.pdf). For further information on the computation and use of standard errors, contact the Decennial Statistical Studies Division at 301-763-4242.

## Chapter 6. User Updates

---

User updates supply data users with additional or corrected information that becomes available after the technical documentation or files are prepared. They are issued as Count Question Resolution Notes, Data Notes, Geography Notes, and Technical Documentation Notes in a numbered series and are available in portable document format (PDF) on our Web site at <http://www.census.gov>.

If you print the documentation, please file the user updates cover sheet behind this notice. If there are technical documentation replacement pages, they should be filed in their proper location and the original pages destroyed.